

# **Development and validation of lifestyle-based models to predict incidence of the most common potentially preventable cancers**

Usher-Smith JA<sup>1</sup>, Sharp SJ<sup>2</sup>, Luben R<sup>3</sup>, Griffin SJ<sup>4</sup>

<sup>1</sup> Clinical Senior Research Associate, The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Box 113 Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK

<sup>2</sup> Senior Statistician, MRC Epidemiology Unit, University of Cambridge, Institute of Metabolic Science, Cambridge, CB2 0QQ, UK

<sup>3</sup> Senior Research Associate, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, 2 Wort's Causeway, Cambridge, CB1 8RN, UK

<sup>4</sup> Professor of General Practice, The Primary Care Unit, Institute of Public Health, University of Cambridge, Box 113 Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK

**Running title:** Lifestyle-based risk score for preventable cancers

**Key words:** Cancer, risk score, lifestyle, prevention, prediction, behaviour

## **Financial support**

This research was funded by a grant from the NIHR School for Primary Care Research (reference number 342). JUS is funded by a Cancer Research UK Prevention Fellowship (C55650/A21464). SJS is supported by the Medical Research Council (unit programme no MC\_ UU\_12015/1). The University of Cambridge has received salary support in respect of SJG from the NHS in the East of England through the Clinical Academic Reserve. EPIC-

Norfolk and RL are supported by the Medical Research Council programme grants (G0401527,G1000143) and Cancer Research UK programme grants (C864/A8257, C864/A14136). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

All researchers were independent of the funding body and the funder had no role in data collection, analysis and interpretation of data; in the writing of the report; or decision to submit the article for publication.

**Correspondence to:**

Dr Juliet Usher-Smith [jau20@medschl.cam.ac.uk](mailto:jau20@medschl.cam.ac.uk)

The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Box 113 Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK

**Competing Interests**

None to declare.

## **ABSTRACT**

### **Background**

Most risk models for cancer are either specific to individual cancers or include complex or predominantly non-modifiable risk factors.

### **Methods**

We developed lifestyle-based models for the five cancers for which the most cases are potentially preventable through lifestyle change in the UK (lung, colorectal, bladder, kidney and oesophageal for men and breast, lung, colorectal, endometrial and kidney for women). We selected lifestyle risk factors from the European Code against Cancer and obtained estimates of relative risks from meta-analyses of observational studies. We used mean values for risk factors from nationally representative samples and mean 10-year estimated absolute risks from routinely available sources. We then assessed the performance of the models in 23,768 participants in the EPIC-Norfolk cohort who had no history of the five selected cancers at baseline.

### **Results**

In men the combined risk model showed good discrimination (AUC: 0.71, 95% CI 0.69-0.73) and calibration. Discrimination was lower in women (AUC: 0.59 95% CI 0.57 – 0.61) but calibration was good. In both sexes the individual models for lung cancer had the highest AUCs (0.83, 95%CI 0.80-0.85 for men and 0.82, 95%CI 0.76-0.87 for women). The lowest AUCs were for breast cancer in women and kidney cancer in men.

### **Conclusions**

The discrimination and calibration of the models are both reasonable, with the discrimination for individual cancers comparable or better than many other published risk models.

### **Impact**

These models could be used to demonstrate the potential impact of lifestyle change on risk of cancer to promote behaviour change.

## INTRODUCTION

Previous research has shown that providing cancer risk information to individuals can improve accuracy of risk perception(1–3), enhance response efficacy(4) and increase intention to have cancer screening(5,6). Additionally, in the only trial assessing the impact of cancer risk tools in primary care on lifestyle behaviours, participants in the intervention group were significantly more likely to report increased daily fruit and vegetable intake and physical activity after six months(7).

Providing individualised estimates of risk of cancer in primary care settings, alongside demonstration of the impact of lifestyle change on that risk, may therefore help motivate change among individuals and complement wider collective approaches to shifting population distributions of behaviour and risk factors. Studies of healthcare professionals and members of the public in the UK have shown that both groups support provision of cancer risk information in primary care(8,9). However, in order to successfully incorporate such risk information into practice, there is a need for risk algorithms which include modifiable risk factors that are either routinely available in electronic medical records or can easily be obtained at new patient registration or within consultations.

The incidence of individual cancers is also low compared to other conditions, such as cardiovascular disease. Consequently, in one study the first reaction of almost all participants to being presented with their 10-year absolute risk of an individual cancer was that it was low and not concerning(9). Providing context for the risk estimates through comparison to other people was also needed. There is, therefore, a need for models that can estimate an individual's combined absolute risk of a number of cancers based on current values of

lifestyle risk factors, and which can be used to calculate relative risk comparing current values with either average or recommended values of the risk factors.

A number of risk models for cancer already exist. However, most are specific to individual cancers(10–13) and while two collections of models exist for multiple cancers, the Q Cancer10 models(14) and the Disease Risk Index(15), to our knowledge, no models that predict risk of multiple cancers together have been published. Additionally, the risk models for individual cancers often include multiple complex risk factors, such as breast density, exposure to asbestos or a past history of colorectal adenomas, or include few modifiable risk factors. Of the 17 models for breast cancer identified in one systematic review(10) for example, less than half included BMI and only one physical activity.

We therefore aimed to develop and validate a lifestyle-based prediction model for the five most common preventable cancers in men and women.

## **MATERIALS AND METHODS**

We developed and validated a risk prediction model in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines(16)

### ***Cancer outcomes***

We included as outcomes the five cancers for which the most cases are potentially preventable through lifestyle change in the UK for men and women from Cancer Research UK data (<http://www.cancerresearchuk.org/health-professional/cancer->

[statistics/risk/preventable-cancers](#)). We excluded melanoma as the evidence to date suggests that risk is influenced by exposure to sun as a child rather than sun protection habits in adulthood (17). For men these are lung, colorectal, bladder, kidney and oesophageal cancer. Together they account for 38% of cancer cases amongst men in the UK and across the five cancers 45,000 cases are estimated to have been preventable in 2011, 25% of all cancer cases in men. For women these are breast, lung, colorectal, endometrial and kidney cancer. Together they account for 61% of cancer cases amongst women in the UK and across the five cancers 48,000 cases are estimated to have been preventable in 2011, 28% of all cancer cases in women.

### ***Development of risk models***

#### ***Risk factors***

We selected lifestyle risk factors for each of these cancers based on the European Code against Cancer 4<sup>th</sup> Edition(18–22). For each cancer we included all risk factors considered convincing or probable by the panels of experts cited within the European Code against Cancer 4<sup>th</sup> Edition, except for: dietary fibre for colorectal cancer in view of the difficulty obtaining reliable self-report measures for fibre intake using a single question; radon levels for lung cancer as this may not always be under the control of the individual; breastfeeding for breast cancer as in most cases this will not be modifiable for the target age range for these risk models (40-79 years); and hormone replacement therapy (HRT) for breast and endometrial cancer because the relationship between HRT use and cancer depends on multiple factors, including type of HRT, route of administration(23), and age at initiation, and the decision to take HRT is a complex decision requiring a balance of risks and benefits and one that should not be made on the basis of future cancer risk alone(24). . We then obtained estimates of the association of each lifestyle factor with each of the relevant cancers from

published meta-analyses of observational studies(25,26,35,36,27–34). For all associations except alcohol with colorectal cancer we assumed a log linear relationship between exposure and risk. To incorporate the increasing evidence that body mass index (BMI) is associated with increased risk of post-menopausal breast cancer but inversely associated with breast cancer before the menopause(30), we included different estimates of the association between BMI and breast cancer in those <50 years of age and those  $\geq 50$  years. No other significant interactions between the risk factors included for the chosen cancers have been reported(30).

#### *Estimates of average values of risk factors*

We estimated average population values of each risk factor in ten year age groups (40-49, 50-59, 60-69 and 70-79 years) in men and women using nationally representative English population surveys. As the latent period between ‘exposure’ to the lifestyle factors and the subsequent increase in cancer risk is not known, we assumed this would be, on average, 10 years as in previous publications(26,37–41). We, therefore, used the Health Survey for England (HSE) 2005 (available from: <https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england>) to obtain population average values for BMI, smoking status, and fruit and vegetable consumption and data from the National Diet and Nutrition survey (NDNS) years 1-4 (2008/12) (available from: <https://discover.ukdataservice.ac.uk/catalogue/?sn=6533>), the closest years to 2005 that included adults up to age 79, for red and processed meat consumption, alcohol intake and physical activity.

Details of the sampling design and methods of both datasets have been described in detail alongside the data (see <https://discover.ukdataservice.ac.uk/catalogue/?sn=6533>). In brief, the HSE is an annual survey designed to measure health and health-related behaviours in a



nationally representative sample of adults and children living in private households in England. In 2005 it also included an additional nationally representative general population sample of adults aged 65 years and over. For this analysis we used data on self-report smoking status, BMI, and portions of fruit and vegetable consumption per day. Each portion of fruit or vegetables was considered 80g based on the British Dietetic Association portion sizes food factsheet (available from:

<https://www.bda.uk.com/foodfacts/portionsizesfoodfactsheet.pdf>) and all analyses were performed with weighting to adjust for non-contact and non-response at both household and individual level.

The NDNS is an annual cross-sectional survey that began in 2008 and covers a representative sample of approximately 1000 people living in private households in the UK per year. For this analysis red meat, processed meat and alcohol consumption (g/day) were obtained from responses to a self-completed food diary. Eight grams of alcohol were estimated as one unit. Physical activity in METs was derived from time spent at moderate or vigorous physical activity, calculated from the Recent physical activity questionnaire (RPAQ)(42). 10 minutes of activity was considered 1 metabolic equivalent (MET)(43). As for the HSE, all analyses were performed with weighting to adjust for non-contact and non-response.

*Estimates of relative risk comparing observed with average values of all the risk factors for separate cancers*

For each of the five cancers, we used the estimates of associations between each risk factor and each cancer to create a formula for the relative risk comparing observed with average values of all the risk factors. For continuous risk factors the relative risk was given by:

$$\text{Relative risk } (\chi) = (\text{risk per unit})^{(\text{observed value} - \text{average value})}$$

For smoking, the other categorical risk factor, the relative risk was given by:

$$\begin{aligned} \text{Relative risk (smoking)} = & (\text{risk for current smokers} \times \text{proportion of population who} \\ & \text{are current smokers}) + (\text{risk for ex-smokers} \times \text{proportion of population who are ex-smokers}) \\ & + (\text{risk for non-smokers} \times \text{proportion of population who are non-smokers}) \end{aligned}$$

Risk factors were assumed to act multiplicatively. The risk of developing endometrial cancer relative to a person with average values of all the risk factors, for example, was calculated by:

$$\begin{aligned} \text{Relative risk} = & (\text{risk per kgm}^{-2} \wedge (\text{observed BMI} - \text{average BMI})) * (\text{risk per MET} \wedge \\ & (\text{observed METs} - \text{average METs})) \end{aligned}$$

#### *Estimates of average 10 year absolute risk for separate cancers*

We then calculated estimates of average 10 year risk for each cancer in men and women in England in the same 10 year age groups from 40 to 79 years using the “Current Probability” method(44). This uses a life-table approach for calculating the risk of developing cancer and takes into account the probability of death from other causes. To obtain the data required for the current probability calculations we used the age- and sex-specific cancer incidence and mortality rates, age- and sex-specific all-cause mortality, and mid-year population size in England during 2015 reported by the Office for National Statistics (available from:

<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatistics/cancerregistrationstatisticsengland> and

<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesanalysisistool>.

*Estimates of 10 year absolute risk with observed values for separate cancers*

To estimate absolute cancer risk over the following 10 years for each separate cancer for an individual with observed values of the risk factors we multiplied the estimates derived above of the relative risk comparing observed with average values of the risk factors by the estimated average 10 year absolute risk for an individual in the same age group and of the same sex as the observed individual, so that:

*10 year absolute risk for an individual = relative risk comparing individual values to average values of lifestyle factors x average 10 year absolute risk for sex and age group.*

*Estimates for the combined cancers*

To estimate the relative risk comparing observed with average values of the risk factors for the five cancers combined we calculated a weighted average of the five cancer-specific relative risks, using the average 10-year estimated absolute risk of developing each cancer in ten year age groups (40 years to 79 years) as weights. The 10 year estimated absolute risk for the combined cancers was calculated by summing the 10 year estimated absolute risks of the five separate cancers, assuming independence of each cancer.

*Estimates of relative risk comparing observed with “recommended” values of all the risk factors*

To allow us to present estimates of the change in risk if individuals followed a “recommended” lifestyle, we used the same method to calculate the relative risk comparing

observed with recommended values of the risk factors. For smoking, BMI, fruit and vegetable consumption and physical activity we used the UK Department of Health guidelines to define these (being a non-smoker, having a body mass index (BMI) of  $25\text{kg/m}^2$ , eating five portions of fruit and vegetables a day, and doing 150 minutes of moderate physical activity per week). For alcohol and red and processed meat consumption which are associated with increased risk, we used zero as our recommended level in line with recommendations from the World Cancer Research Fund (<https://www.wcrf-uk.org/>). This decision was made to avoid appearing to encourage consumption of red or processed meat or alcohol among those consuming small amounts.

### ***Validation of risk model***

We externally validated the model in the EPIC-Norfolk cohort(45). This includes 25,639 men and women who were recruited between March 1993 and December 1997 from 27 practices across Norfolk and were aged between 45-74 years old at the time of recruitment.

Participants were extensively phenotyped at baseline. Incident cases of cancer are recorded through linkage to the cancer registries. Smoking status and alcohol consumption were assessed using single questions Alcohol consumption in grams was estimated as the total units of drinks consumed in a week multiplied by eight. Fruit, vegetable, red meat and processed meat consumption were estimated from responses to a previously validated food frequency questionnaire and seven day food diary respectively(46): Each portion of fruit or vegetables was considered 80g and each portion of red or processed meat 90g, based on the British Dietetic Association portion sizes food factsheet (available from:

<https://www.bda.uk.com/foodfacts/portionsizesfoodfactsheet.pdf>). Physical activity was computed from the average number of hours per week that participants reported cycling or

doing “other physical activity such as keep fit, aerobics, swimming or jogging”. As above, 10 minutes of activity was considered 1 metabolic equivalent (MET)(43).

We assessed the performance of the risk model in 23,768 participants (12,828 women and 10,940 men) in the EPIC-Norfolk cohort who had at least 10 year follow-up, data for all risk factors and no previous history or diagnosis of any of the chosen cancers at baseline.

Participants with prevalent and incident diagnoses of each cancer were identified through linkage to the National Cancer Registry. A full list of the ICD9 and ICD10 codes used for each cancer are in Supplementary Table 1. We truncated continuous variables at the 95<sup>th</sup> centile. We treated the outcome as a binary variable (developed one or more of the five cancers or did not develop any of the five cancers). For the primary analysis we assessed discrimination by plotting receiver operating characteristic (ROC) curves and calculating the area under the ROC curve (AUC). We assessed calibration graphically by plotting the observed risk (i.e. percentage of individuals who developed cancer) within each decile of predicted risk and calculated overall observed to expected ratios.

All analyses were conducted using STATA/SE version 14.2(47)

## **RESULTS**

### **Development of risk model**

Details of the lifestyle risk factors included, units for comparison, relative risks and their source are given in Table 1. Smoking is associated with the highest relative risk for five of the seven cancers considered across the two sexes, followed by BMI. The sites associated with the greatest number of lifestyle risk factors are colorectal and oesophageal cancer. The

estimated mean 10-year risk for each of the cancers is provided in Supplementary Tables 2a and b. In men, colorectal cancer has the highest 10-year estimated risk of the five cancers between 20 and 59 years and lung cancer the highest above age 60 years. In women, breast cancer has the highest 10-year estimated risk at all ages, with colorectal cancer the second highest up to age 60 years and lung cancer above age 60 years. The mean values used in the calculations are reported in Supplementary Table 3. Across all age groups there were more former smokers amongst men than women and men consumed more alcohol, more red and processed meat and less fruit and, except for those aged 70-79 years, less vegetables than women. In both sexes, alcohol consumption, vegetable consumption and physical activity reduced with age while fruit consumption increased up to age 60-69.

### **Performance of risk model**

From the 25,059 participants within the EPIC-Norfolk cohort who were aged 40-79 years at baseline with no existing diagnosis of any of the cancers of interest and who had 10 year follow-up, data for all risk factors were available for 23,768 (94.8%). Among these participants, 432 (3.95%) men and 647 (5.0%) women developed at least one of the cancers during the 10 year follow-up (Tables 2 and 3). Compared to those who did not develop at least one of the cancers, those who did were on average older and more likely to be female, have a higher BMI, report less physical activity and be a current smoker. Differences in red meat, processed meat, fruit and vegetable consumption were small.

The mean relative risk compared with the ‘recommended’ lifestyle was 1.76 (SD 0.94, range 0.72 to 8.0). In men there was good discrimination with an AUC for the combined model of 0.71 (95% CI 0.69-0.73) (Figure 1a). There was also reasonable agreement between the predicted absolute 10-year risk and the observed risk (Figure 2a) although overall the risk

model underestimated risk with an observed:expected ratio of 1.34 (95% CI 1.04-1.73). Discrimination was less good in women with an AUC of 0.59 (95% CI 0.57 – 0.61) (Figure 1b). Overall calibration was better in women (Figure 2b) with the overall observed to expected ratio 1.08 (95% CI 0.90-1.30) but at higher risks the model overestimated risk.

Figures 3a and b show the AUC for the five individual cancers as well as the combined model for men and women respectively. In both sexes the models for lung cancer had the highest AUCs (0.80 (95% CI 0.77-0.83) for men and 0.82 (95% CI 0.76-0.87) for women). The lowest AUCs were for breast and endometrial cancer in women and kidney cancer in men.

## **DISCUSSION**

### **Key findings**

We have developed and validated models in men and women for prediction of the individual or combined absolute risk of developing one or more of the five cancers for which the most cases are potentially preventable through lifestyle change. The models can also be used to present relative risks compared with either an average or a recommended lifestyle. The models include information about established lifestyle risk factors in a format that is readily obtainable from individuals or their medical records without the need for laboratory tests. The combined models had good discrimination in men (AUC 0.71) and reasonable discrimination in women (AUC 0.59). The discrimination for the individual cancers ranged from very good for lung cancer (AUC over 0.8 in both men and women), to poor for breast cancer (AUC 0.56). Overall calibration, as assessed graphically, was reasonable. The models could be used to identify those most likely to benefit from lifestyle interventions and to demonstrate the

impact of change to individuals and comparison of their risk to others to contribute to decisions to change behaviour.

### **Strengths and Limitations**

To our knowledge these are the first risk models incorporating only modifiable factors alongside age and sex that have been developed for multiple cancers in a UK population. Particular strengths include the use of the European Code against Cancer 4<sup>th</sup> Edition to identify the risk factors associated with each of the individual cancers; the use of data from systematic reviews to obtain estimates of the relative risk for each lifestyle factor and each of the relevant cancers; the use of nationally representative datasets to obtain average values for each risk factor; and assessment of the performance of the models in a large population-based UK cohort. There are, however, a number of limitations. Firstly, except for the association between alcohol and colorectal cancer, we assumed a log-linear relationship between exposure and relative risk. This is supported by the absence of significant non-linearity reported in many of the meta-analyses but may have influenced estimates of relative risk for extreme values of each risk factor.

Estimating the absolute 10-year risk for an individual required us to estimate the average 10-year risk at varying ages for each of the cancers. We did this using the current probability method. This involves calculating the number of cases that would occur within each age band on the basis of the person-years at risk and age-specific incidence rate(44). This has the advantage over cumulative risk estimates in that it takes into account competing risks. However, when it is calculated using routine incidence data, which includes multiple primary cancers, such as the data from ONS which we used, the method is actually estimating the average number of primary cancers per person rather than the probability of a person getting



cancer. As a result the method tends to overestimate lifetime risk of getting cancer for individual cancers. However, the difference between the estimates obtained from this method and the ‘gold standard’ estimates have been shown to be small(48).

Advantages of using the EPIC-Norfolk cohort for validation of the models include the comprehensive phenotyping, completeness of data, and linkage to national cancer registries. However, although the cohort at baseline was representative of the Health Survey for England population for age, sex and BMI, it had fewer current smokers(45) and participants were recruited from only one geographical region in the East of England. The models therefore need to be assessed in other populations before inferences can be made about model generalizability(49).

### **Comparison with existing literature**

As we aimed to develop models that could be used within routine practice, we have not included all factors that are predictive of each cancer. We also specifically sought to include variables in a way that they could be collected easily and quickly. Had we included greater detail for variables, for example using pack years of smoking rather than a categorical smoking status variable, model performance may have been improved. Despite this, the discrimination for the individual cancers are comparable or better than many other published risk models. For example: the AUCs of 0.66 and 0.68 for colorectal cancer in men and women respectively are comparable with published models that also include family history and more complex variables(50); the AUC of over 0.80 for lung cancer in both men and women is better than the range of 0.61 to 0.81 reported in external validation of nine models, all of which included age and smoking behaviour(51); and the AUC of 0.74 for bladder cancer in men is better than the only other published model(52) which includes smoking and

exposure to diesel, aromatic amines, dry cleaning fluids, radioactive materials and arsenic and had an AUC of 0.70 (95% CI 0.67-0.73) in split-sample validation.

With an AUC of 0.55, the model for breast cancer has the lowest discrimination. This is consistent with the literature in which models incorporating a combination of age, age at menarche, first live birth and menopause, breast biopsy and family history of breast cancer still only have AUCs between 0.59 and 0.67 in population based cohorts(53,54).

Furthermore, a recent extension of the Rosner-Colditz prediction model incorporating 22 risk factors had an AUC of only 0.65 in split-sample validation(55). As described previously, the relatively weak predictive ability of these models may arise because risk factors with large associations such as mutations in the BRCA1 or BRCA2 genes have low prevalence, and risk factors such as early age at menarche or late age at first birth are common amongst women who never go on to develop breast cancer and have only modest associations with risk.

The AUCs for endometrial cancer (AUC 0.61) and oesophageal cancer (AUC 0.65) were also lower than other published models with AUCs of 0.68 (95% CI 0.66-0.70) in external validation(56) and 0.77 (0.68-0.85) in cross-validation(57) for endometrial cancer and 0.79 (95% CI 0.75-0.83) in cross-validation(52) and 0.75 (95% CI 0.66-0.84) in 10-fold cross-validation(58) for oesophageal cancer. This may reflect the absence from our models of hormonal factors for endometrial cancer and reflux symptoms or non-steroidal anti-inflammatory drugs for oesophageal cancer. To our knowledge no other risk models have been developed for kidney cancer alone.

The discrimination of the combined model for the five most common preventable cancers in men (AUC 0.71) is also comparable with the Healthy Heart Score, a lifestyle-based

prediction model for cardiovascular disease(59). The Healthy Heart Score includes age, smoking status, BMI, physical activity, alcohol and a composite diet score incorporating fibre, fruit and vegetables, nuts, sugar-sweetened beverages and red and processed meats. In split-sample validation within large US based cohorts it had a Harrell's C-index of 0.77 (95% CI 0.76-0.79) in men and 0.72 (95% CI 0.71-0.74) in women. The performance of the combined model for the five most common preventable cancers in women (AUC 0.59) was, however, substantially poorer. This is in part due to the poor performance of the risk score for breast cancer which is the most common cancer in women and so has the greatest weighting within the combined model. As mentioned above, hormonal factors are also known to influence risk of cancer in women and are not included in our models.

### **Implications for clinicians, researchers and policy makers**

The risk models developed here are applicable to individuals aged 40 to 79 years and enable presentation of information about the impact of lifestyle on future risk of cancer. By focusing on the most common preventable cancers and modifiable risk factors which can be easily obtained without the need for laboratory or imaging tests, they could be used as the first step in a multi-stage risk stratification programme to identify those most likely to benefit from lifestyle interventions and to motivate individuals to change their behaviour. For example, a 65 year old man who weighs 80kg, is 1.7m tall, is a current smoker, drinks four units of alcohol per day, eats three portions of red meat and two portions of processed meat per week and one portion of fruit and vegetables per day, and does two hours of moderate physical activity per week has an estimated 10-year risk of 8%. If he lost 5kg of weight, quit smoking, reduced alcohol to one unit per day and red and processed meat to two portions per week, increased fruit and vegetables to five portions per day and did three hours of physical activity per week, his risk would reduce to 4%. Previous research has shown that public awareness of

the impact of such lifestyle changes on cancer risk is low: only 3% of people are aware that being overweight can increase their risk of cancer; less than a third are aware that physical activity could help reduce risk(60–63); and one in seven people believe that lifetime risk of cancer is unmodifiable(64). Although communicating personalised risk information on its own is unlikely to lead to much sustained behaviour change(65) using these risk models to illustrate the impact of lifestyle change and allow individuals to compare their risk to average people of their age and sex and the risk if they followed a recommended lifestyle may help motivate change at an individual level when combined with other established behaviour change techniques(66). This would then complement wider collective approaches to shifting population distributions of behaviour and risk factors.

This risk assessment could be conducted within existing healthcare and prevention services(8) or made available online. Further research is now needed to assess the performances of the models in other populations, develop a user-friendly interface in which these models can be incorporated into clinical practice, and implementation studies quantifying the potential benefits and harms of providing such information.

## **ACKNOWLEDGEMENTS**

The authors would like to acknowledge the contribution of the staff and participants of the EPIC-Norfolk Study and Professor Kay-Tee Khaw for helpful comments on the manuscript.

## **Contributors**

JUS and SJG conceived and designed the study. JUS carried out the statistical analysis. SJS provided statistical advice. RL provided advice on the EPIC-Norfolk cohort. All authors interpreted the data. JUS drafted the manuscript and all authors critically revised the manuscript. JUS is the study guarantor.

## REFERENCES

1. Emmons KM, Wong M, Puleo E, Weinstein N, Fletcher R, Colditz G. Tailored computer-based cancer risk communication: correcting colorectal cancer risk perception. *J Health Commun.* 2004;9:127–41.
2. Weinstein ND, Atwood K, Puleo E, Fletcher R, Colditz G, Emmons KM. Colon Cancer: Risk Perceptions and Risk Communication. *J Health Commun.* 2004;9:53–65.
3. Wang C, Sen A, Ruffin MT, Nease DE, Gramling R, Acheson LS, et al. Family history assessment: impact on disease risk perceptions. *Am J Prev Med.* 2012;43:392–8.
4. Cameron LD, Marteau TM, Brown PM, Klein WM, Sherman KA. Communication strategies for enhancing understanding of the behavioral implications of genetic and biomarker tests for disease risk: the role of coherence. *J Behav Med.* 2011/06/24. 2012;35:286–98.
5. Schroy PC, Emmons KM, Peters E, Glick JT, Robinson PA, Lydotes MA, et al. Aid-assisted decision making and colorectal cancer screening: a randomized controlled trial. *Am J Prev Med.* 2012;43:573–83.
6. Schroy PC, Emmons K, Peters E, Glick JT, Robinson PA, Lydotes MA, et al. The impact of a novel computer-based decision aid on shared decision making for colorectal cancer screening: a randomized trial. *Med Decis Mak.* 2011;31:93–107.
7. Ruffin MT, Nease DE, Sen A, Pace WD, Wang C, Acheson LS, et al. Effect of preventive messages tailored to family history on health behaviors: the Family Healthware Impact Trial. *Ann Fam Med.* 2011;9:3–11.
8. Usher-smith JA, Silarova B, Ward A, Youell J, Muir KR. Incorporating cancer risk information into general practice : a qualitative study using focus groups with healthcare professionals. *BJGP.* 2017;67:e218–26.
9. Usher-Smith JA, Silarova B, Lophatananon A, Duschinsky R, Campbell J, Warcaba J,

- et al. Responses to provision of personalised cancer risk information: a qualitative interview study with members of the public. *BMC Public Health*. 2017;17:977.
10. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. *Breast Cancer Res Treat*. 2012;132:365–77.
  11. Cassidy A, Duffy SW, Myles JP, Liloglou T, Field JK. Lung cancer risk prediction: a tool for early detection. *Int J cancer*. 2007;120:1–6.
  12. Win AK, Macinnis RJ, Hopper JL, Jenkins M a. Risk prediction models for colorectal cancer: a review. *Cancer Epidemiol Biomarkers Prev*. 2012;21:398–410.
  13. Ma GK, Ladabaum U. Personalizing Colorectal Cancer Screening: A Systematic Review of Models to Predict Risk of Colorectal Neoplasia. *Clin Gastroenterol Hepatol*. 2014;12:1624–34.e1.
  14. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithms to estimate future risk of common cancers in men and women: prospective cohort study. *BMJ Open*. 2015;5:e007825.
  15. Colditz GA, Atwood KA, Emmons K, Monson RR, Willett WC, Trichopoulos D, et al. Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. Risk Index Working Group, Harvard Center for Cancer Prevention. *Cancer Causes Control*. 2000;11:477–88.
  16. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med*. 2015;162:55–63.
  17. Usher-Smith JA, Emery J, Kassianos AP, Walter FM. Risk Prediction Models for Melanoma: A Systematic Review. *Cancer Epidemiol biomarkers Prev*. 2014;23:1450–63.

18. Leon M, Peruga A, McNeill A, Kralikova E, Guha N, Minozzi S, et al. European Code against Cancer, 4th Edition: Tobacco and cancer. *Cancer Epidemiol.* 2015;39S:S20–33.
19. Scoccianti C, Cecchini M, Anderson AS, Berrino F, Boutron-Ruault MC, Espina C, et al. European Code against Cancer 4th Edition: Alcohol drinking and cancer. *Cancer Epidemiol.* Elsevier Ltd; 2015;39S:S67–74.
20. Norat T, Scoccianti C, Boutron-Ruault MC, Anderson AS, Berrino F, Cecchini M, et al. European Code against Cancer 4th Edition: Diet and cancer. *Cancer Epidemiol.* Elsevier Ltd; 2015;39S:S56–66.
21. Anderson AS, Key TJ, Norat T, Scoccianti C, Cecchini M, Berrino F, et al. European Code against Cancer 4th Edition: Obesity, body fatness and cancer. *Cancer Epidemiol.* Elsevier Ltd; 2015;39:S34–45.
22. Leitzmann M, Powers H, Anderson AS, Scoccianti C, Berrino F, Boutron-Ruault MC, et al. European Code against Cancer 4th Edition: Physical activity and cancer. *Cancer Epidemiol.* Elsevier Ltd; 2015;39:S46–55.
23. Wang K, Li F, Chen L, Lai Y-M, Zhang X, Li H-Y. Change in risk of breast cancer after receiving hormone replacement therapy by considering effect-modifiers: a systematic review and dose-response meta-analysis of prospective studies. *Oncotarget.* 2017;8:81109–24.
24. National Institute for Health and Care Excellence. Menopause: diagnosis and management. CG23. 2015.
25. Wu Y, Zhang D, Kang S. Physical activity and risk of breast cancer: A meta-analysis of prospective studies. *Breast Cancer Res Treat.* 2013;137:869–82.
26. Parkin D. 9. Cancers attributable to inadequate physical activity in the UK in 2010. *Br J Cancer.* Nature Publishing Group; 2011;105:S38–41.



27. Keum N, Ju W, Lee DH, Ding EL, Hsieh CC, Goodman JE, et al. Leisure-time physical activity and endometrial cancer risk: Dose-response meta-analysis of epidemiological studies. *Int J Cancer*. 2014;135:682–94.
28. Gandini S, Botteri E, Iodice S, Boniol M, Lowenfels AB, Maisonneuve P, et al. Tobacco smoking and cancer: A meta-analysis. *Int J Cancer*. 2008;122:155–64.
29. van Osch FHM, Jochems SHJ, van Schooten F-J, Bryan RT, Zeegers MPA. Quantified relations between exposure to tobacco smoking and bladder cancer risk: a meta-analysis of 89 observational studies. *Int J Epidemiol*. 2016;45:dyw044.
30. WRCF. Continuous update project (CUP) systematic literature reviews. 2017; Available from: [https://www.wcrf.org/sites/default/files/CUP Summary Report May17.pdf](https://www.wcrf.org/sites/default/files/CUP%20Summary%20Report%20May17.pdf)
31. Chen WY, Willett WC. Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk. *JAMA J Am Med Assoc*. 2012;307:666.
32. Fedirko V, Tramacere I, Bagnardi V, Rota M, Scotti L, Islami F, et al. Alcohol drinking and colorectal cancer risk: An overall and dose-Response meta-analysis of published studies. *Ann Oncol*. 2011;22:1958–72.
33. Wang F, Xu Y. Body mass index and risk of renal cell cancer: A dose-response meta-analysis of published cohort studies. *Int J Cancer*. 2014;135:1673–86.
34. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*. 2008;371:569–78.
35. World Cancer Research Fund/American Institute for Cancer Research. Food, nutrition, physical activity and the prevention of cancer: a global perspective. 2007.
36. Soerjomataram I, Oomen D, Lemmens V, Oenema A, Benetou V, Trichopoulou A, et al. Increased consumption of fruit and vegetables and future cancer incidence in

- selected European countries. *Eur J Cancer*. Elsevier Ltd; 2010;46:2563–80.
37. Parkin D. 8. Cancers attributable to overweight and obesity in the UK in 2010. *Br J Cancer*. Nature Publishing Group; 2011;105:S34–7.
  38. Parkin D. 3. Cancers attributable to consumption of alcohol in the UK in 2010. *Br J Cancer*. Nature Publishing Group; 2011;105:S14–8.
  39. Parkin D, Boyd L. 4. Cancers attributable to dietary factors in the UK in 2010 I. Low consumption of fruit and vegetables. *Br J Cancer*. Nature Publishing Group; 2011;105:S19–23.
  40. Parkin D. 5. Cancers attributable to dietary factors in the UK in 2010 II. Meat consumption. *Br J Cancer*. Nature Publishing Group; 2011;105:S24–6.
  41. Parkin D. 2. Tobacco-attributable cancer burden in the UK in 2010. *Br J Cancer*. Nature Publishing Group; 2011;105:S6–13.
  42. Wareham NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, Day NE. Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire. *Int J Epidemiol*. 2002;31:168–74.
  43. Ainsworth B, Haskell W, Leon A, Jacobs D, Montoye H, Sallis J, et al. Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sport Exerc*. 1993;25:71–80.
  44. Estève J, Benhamou E, Raymond L. Statistical methods in cancer research. Volume IV. Descriptive epidemiology. *IARC Sci Publ*. 1994;1–302.
  45. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A, et al. EPIC-Norfolk: study design and characteristics of the cohort. *European Prospective Investigation of Cancer*. *Br J Cancer*. 1999;80 Suppl 1:95–103.
  46. Bingham SA, Welch AA, McTaggart A, Mulligan AA, Runswick SA, Luben R, et al. Nutritional methods in the European Prospective Investigation of Cancer in Norfolk.

- Public Health Nutr. 2001;4:847–58.
47. StataCorp. Stata Statistical Software: Release 13. TX StataCorp LP. 2013;
  48. Sasieni PD, Shelton J, Ormiston-Smith N, Thomson CS, Silcocks PB. What is the lifetime risk of developing cancer?: the effect of adjusting for multiple primaries. *Br J Cancer*. Nature Publishing Group; 2011;105:460–5.
  49. Kundu S, Mazumdar M, Ferket B. Impact of correlation of predictors on discrimination of risk models in development and external populations. *BMC Med Res Methodol*. 2017;17:63.
  50. Usher-Smith JA, Walter FM, Emery J, Win AK, Griffin SJ. Risk prediction models for colorectal cancer: a systematic review. *Cancer Prev Res (Phila)*. 2015;
  51. ten Haaf K, Jeon J, Tammemägi MC, Han SS, Kong CY, Plevritis SK, et al. Risk prediction models for selection of lung cancer screening candidates: A retrospective validation study. *PLoS Med*. 2017;14:1–24.
  52. Wu X, Lin J, Grossman HB, Huang M, Gu J, Etzel CJ, et al. Projecting Individualized Probabilities of Developing Bladder Cancer in White Individuals. *J Clin Oncol*. 2007;25:4974–81.
  53. Costantino JP, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst*. 1999;91:1541–8.
  54. Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *Am J Epidemiol*. 2000;152:950–64.
  55. Glynn RJ, Colditz GA, Tamimi RM, Chen WY, Hankinson SE, Willett WW, et al. Extensions of the Rosner-Colditz breast cancer prediction model to include older women and type-specific predicted risk. *Breast Cancer Res Treat*. Springer US;

- 2017;165:215–23.
56. Pfeiffer RM, Park Y, Kreimer AR, Lacey J V., Pee D, Greenlee RT, et al. Risk Prediction for Breast, Endometrial, and Ovarian Cancer in White Women Aged 50 y or Older: Derivation and Validation from Population-Based Cohort Studies. *PLoS Med.* 2013;10.
  57. Hüsing A, Dossus L, Ferrari P, Tjønneland A, Hansen L, Fagherazzi G, et al. An epidemiological model for prediction of endometrial cancer risk in Europe. *Eur J Epidemiol.* 2016;31:51–60.
  58. Thrift AP, Kendall BJ, Pandeya N, Whiteman DC. A Model to Determine Absolute Risk for Esophageal Adenocarcinoma. *Clin Gastroenterol Hepatol.* W.B. Saunders; 2013;11:138–144.e2.
  59. Chiuve SE, Cook NR, Shay CM, Rexrode KM, Albert CM, Manson JAE, et al. Lifestyle-based prediction model for the prevention of CVD: The healthy heart score. *J Am Heart Assoc.* 2014;3:1–11.
  60. Cancer Research UK. Perceptions of Risk Survey 2008 : Key Findings. 2008; Available from:  
[www.cancerresearchuk.org/sites/default/files/perceptions\\_of\\_risk\\_survey.pdf](http://www.cancerresearchuk.org/sites/default/files/perceptions_of_risk_survey.pdf)
  61. Grunfeld EA, Ramirez AJ, Hunter MS, Richards MA. Women’s knowledge and beliefs regarding breast cancer. *Br J Cancer.* 2002;86:1373–8.
  62. Redeker C, Wardle J, Wilder D, Hiom S, Miles A. The launch of Cancer Research UK’s “Reduce the Risk” campaign: baseline measurements of public awareness of cancer risk factors in 2004. *Eur J Cancer.* 2009;45:827–36.
  63. Wardle J, Waller J, Brunswick N, Jarvis MJ. Awareness of risk factors for cancer among British adults. *Public Health.* 2001;115:173–4.
  64. Ryan AM, Cushen S, Schellekens H, Bhuachalla EN, Burns L, Kenny U, et al. Poor

- Awareness of Risk Factors for Cancer in Irish Adults: Results of a Large Survey and Review of the Literature. *Oncologist*. 2015;20:372–8.
65. French DP, Cameron E, Benton JS, Deaton C, Harvie M. Can Communicating Personalised Disease Risk Promote Healthy Behaviour Change? A Systematic Review of Systematic Reviews. *Ann Behav Med. Annals of Behavioral Medicine*; 2017;
66. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions. *Ann Behav Med*. 2013;46:81–95.

**Table 1. Estimates of the association between risk factors and individual cancers**

<b>Risk factor</b>	<b>Unit / comparison</b>	<b>Relative risk</b>	
<b>Physical activity(22)</b>	1 MET-hour per week	Breast(25) – 0.9970 Colorectal(26) –0.9940 Endometrial(27) – 0.9933	
<b>Smoking(18)</b>	Current and former smokers compared with non-smokers	<i>Current smokers</i> Lung(28) – 8.96 Colorectal(28) - 1.2 Oesophageal(28) – 2.5 Kidney(28) – 1.52 Bladder(29) – 3.14	<i>Former smokers</i> Lung(28) – 3.85 Colorectal(28) - 1.2 Oesophageal(28) – 2.03 Kidney(28) – 1.25 Bladder(29) – 1.83
<b>Red meat(20)</b>	1 gram per day	Colorectal(30) – 1.0016	
<b>Processed meat(20)</b>	1 gram per day	Colorectal(30) – 1.0033	
<b>Alcohol(19)</b>	1 gram per day	Breast(31) – 1.0068 Colorectal(32) – $\ln(RR) = 0.006992 \times \text{g/day} - 0.00001 \times \text{g/day}^2$ Oesophageal(33) – 1.0129	
<b>BMI(25)</b>	1 kg/m <sup>2</sup>	Breast (postmenopausal*) (30) – 1.0229 Breast (premenopausal*) (30) - 0.9856 Colorectal(33) – 1.030 Oesophageal(34) – 1.087 Kidney(33) – 1.04 Endometrial(30) – 1.034	
<b>Fruit(20)</b>	1 gram per day	Oesophageal(35) – 0.994 Lung(36) – 0.99	
<b>Vegetables(20)</b>	1 gram per day	Oesophageal(35) – 0.9972	

\* Defined as below 50 years or 50 years and over such that premenopausal = age < 50 years and post-menopausal = age ≥ 50 years

**Table 2. Demographic and lifestyle characteristics of the EPIC-Norfolk cohort**

	Validation cohort	
	No incident cancer <i>n</i> = 22 689	Incident cancer <i>n</i> = 1 079
Age (years)		
Mean (SD)	58.9 (9.3)	62.9 (8.6)
40-49 (%)	22.2	8.7
50-59 (%)	32.2	27.2
59-60 (%)	30.7	37.4
69-70 (%)	15.0	26.8
Sex		
Female (%)	53.7	60.0
Male (%)	46.3	40.0
BMI (kg/m <sup>2</sup> )		
Mean (SD)	26.3 (3.9)	26.9 (4.3)
Smoking status		
Never (%)	42.39	43.0
Former (%)	46.13	42.5
Current (%)	11.5	14.5
Alcohol intake (units per day)		
Mean (SD)	7.5 (8.2)	7.2 (8.2)
Physical activity (MET-h/d)		
Mean (SD)	9.3 (13.2)	8.4 (13.0)
Red meat consumption (g/week)		
Mean (SD)	39.3 (25.3)	39.8 (25.0)
Processed meat consumption (g/week)		
Mean (SD)	17.9 (12.9)	18.0 (12.9)
Fruit consumption (g/day)		
Mean (SD)	251.2 (149.6)	254.7 (154.2)
Vegetable consumption (g/day)		
Mean (SD)	251.2 (107.11)	252.3 (105.6)

**Table 3. Incident cases of cancer (n, %) within the EPIC-Norfolk validation cohort**

Cancer	Men ( <i>n</i> =10,940)	Women ( <i>n</i> =12,828)
Lung	142 (1.30)	54 (0.42)
Colorectal	184 (1.68)	138 (1.08)
Kidney	28 (0.26)	16 (0.12)
Breast	----	367 (2.86)
Bladder	47 (0.43)	----
Endometrial	----	84 (0.65)
Oesophageal	35 (0.32)	----



## **FIGURE LEGENDS**

Figure 1. Receiver operating characteristic curves for the combined model in a) men; and b) women.

Figure 2. Calibration plots of observed risk in the EPIC-Norfolk cohort within each decile of predicted risk in a) men; and b) women. Error bars represent the 95% confidence intervals.

Figure 3. Area under the receiver operating characteristic curves (AUC) for the five models of individual cancers and the combined model for a) men; and b) women.

Figure 1

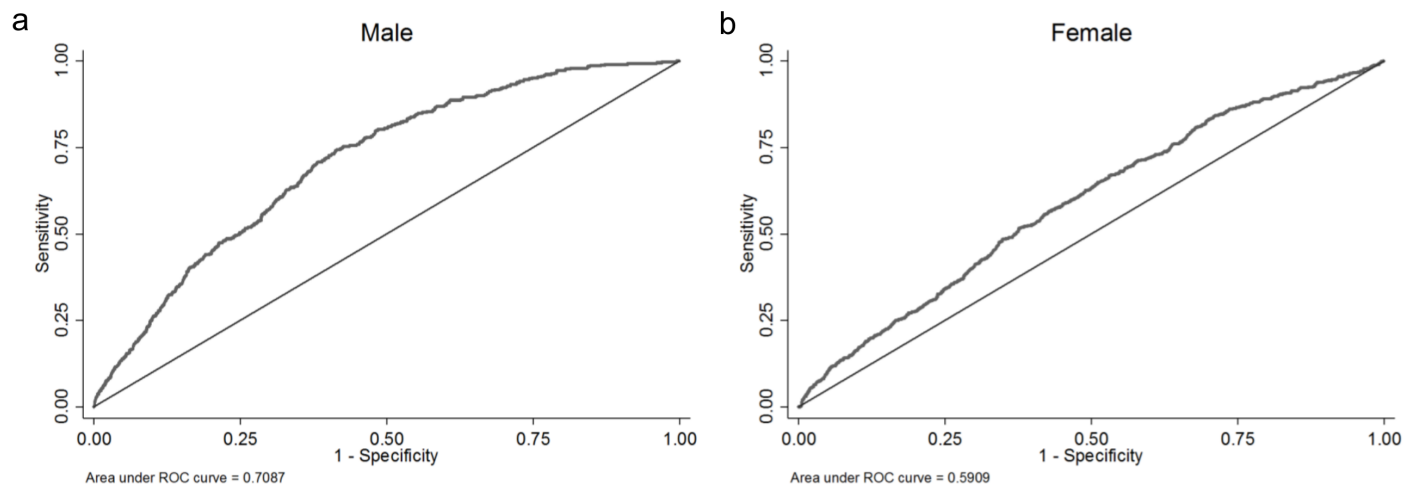


Figure 2

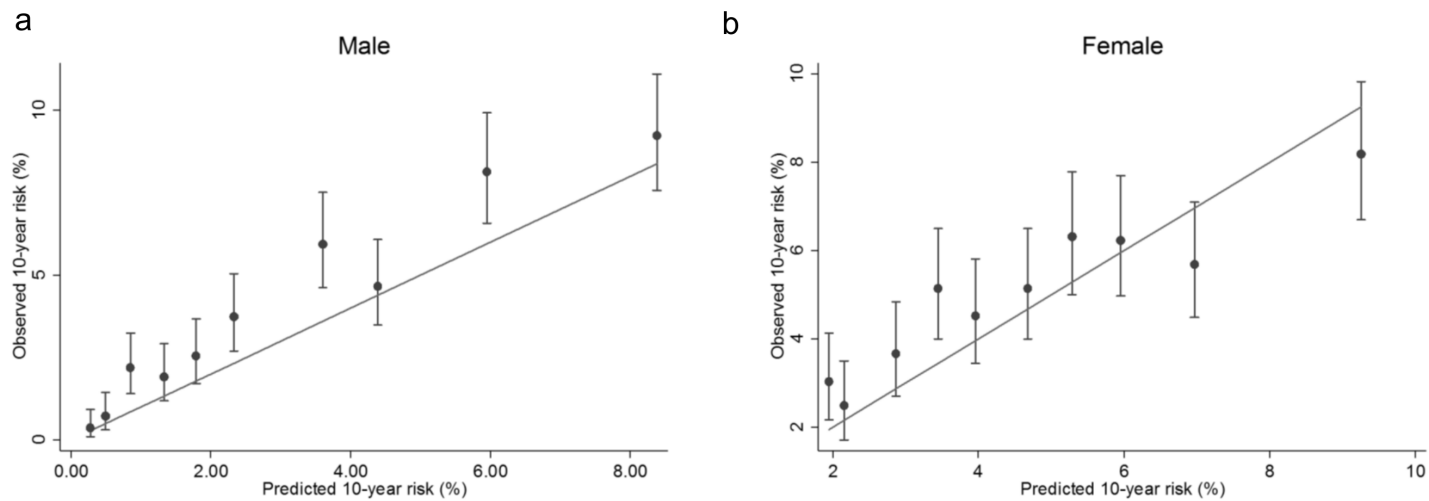


Figure 3

